Insight 04 December 2024 | 11 min read

AI Watch: Global regulatory tracker - G7

The G7's AI regulations mandate Member States' compliance with international human rights law and relevant international frameworks.

Laws/Regulations directly regulating AI (the “AI Regulations”)

The G7 nations have progressed the Hiroshima AI Process Comprehensive Policy Framework, which consists of four pillars:

(i) the International Guiding Principles for Organizations Developing Advanced AI Systems (the "Guiding Principles");[1]

(ii) the International Code of Conduct for Organizations Developing Advanced AI Systems (the "Code of Conduct")[2] designed to supplement the Guiding Principles and provide voluntary guidance to organizations developing Advanced AI systems;

(iii) analysis of priority risks, challenges and opportunities of generative AI; and

(iv) project-based cooperation in support of the development of responsible AI tools and best practices.

Neither the Guiding Principles nor the Code of Conduct are legally binding, yet both pieces of guidance will likely exert strong political influence internationally.

Status of the AI Regulations

The G7 lacks the ability to pass laws regarding AI or its implementation. Nevertheless, the G7's AI Regulations do specify that its members must abide by their obligations under international human rights law, while private sector activities should be in line with international frameworks such as the United Nations Guiding Principles on Business and Human Rights and the OECD Guidelines of Multinational Enterprises.

The Guiding Principles were proposed in draft form on October 30, 2023 and are expected to be regularly reviewed and updated. This will involve multiple consultations before the Guiding Principles are finalized.

The International Code of Conduct also remains in draft form as proposed on October 30, 2023. Timing of the consultation and finalization process remains uncertain.[3]

The G7 competition authorities and the European Commission met in Italy on October 3-4, 2024, to address competition challenges in digital markets posed by new technologies. Their Joint Statement[4] and AI working group's Discussion Paper[5] outline risks for businesses and strategies to ensure competition, protect innovation, and promote responsible AI practices.

Other laws affecting AI

The G7's Guiding Principles and Code of Conduct build on existing OECD AI Principles and are intended to inform and spearhead the national regulatory regimes implemented by the G7 nations as part of a fit-for-purpose global governance charter on AI.

In addition, there are various laws and frameworks that do not directly seek to regulate AI, but may affect the development or use of AI in the G7. For example:

- International human rights law continues to apply to the G7 states to ensure that human rights are fully respected and protected
- Private sector activities of all AI actors should comply with international frameworks such as the United Nations Guiding Principles on Business and Human Rights and the OECD Guidelines for Multinational Enterprises

Definition of “AI”

The Guiding Principles and Code of Conduct do not establish an independent definition of "AI." Instead, both pieces of guidance build on the OECD AI Principles that adopt the following definitions:[6]

- "**AI system**" means "a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment."
- "**AI actors**" means "those who play an active role in the AI system lifecycle, including organizations and individuals that deploy or operate AI."
- "**AI system lifecycle**" involves the following phases: "i) 'design, data and models,' which is a context-dependent sequence encompassing planning and design, data collection and processing, as well as model building; ii) 'verification and validation,' iii) 'deployment,' and iv) 'operation and monitoring.' These phases often take place in an iterative manner and are not necessarily sequential. The decision to retire an AI system from operation may occur at any point during the operation and monitoring phase."

The Code of Conduct and the Guiding Principles additionally define the following term:

- "**Advanced AI systems**" to mean "the most advanced foundation models and generative AI systems."

Territorial scope

The Guiding Principles and Code of Conduct do not territorially confine the concepts of AI actors or Advanced AI systems. The G7 members, namely Canada, France, Germany, Italy, Japan, the UK, the US, and the EU have called on AI actors in their respective states to follow the Guiding Principles and have called on organizations to follow the Code of Conduct in line with a risk-based approach while national governments develop more detailed governance and regulatory regimes.

Sectoral scope

The Guiding Principles and Code of Conduct are not sector-specific.

The Guiding Principles apply to all AI actors (i.e., including both individuals and organizations) involved in the design, development, deployment and use of Advanced AI systems.

The Code of Conduct applies to all organizations that are developing Advanced AI systems.

Such organizations may include entities from academia, civil society, the private sector and the public sector.[7]

Compliance roles

Organizations involved in the design, development, deployment and use of Advanced AI Systems are expected to abide by the G7's AI Regulations. All AI actors should also comply with the Guiding Principles.

Core issues that the AI Regulations seek to address

The G7's AI Regulations seek to promote safe, secure, and trustworthy AI worldwide and provide practical guidance for organizations developing and using foundation models and generative AI systems. The G7's AI Regulations actively seek to prevent organizations from developing or deploying Advanced AI systems that are considered "not acceptable" – namely Advanced AI systems that undermine democratic values, are particularly harmful to individuals or communities, facilitate terrorism, enable criminal misuse, or pose substantial risks to safety, security, and human rights.[8]

Risk categorization

AI is not explicitly categorized according to risk in the G7's AI Regulations. However, the Code of Conduct highlights various risks that should be particularly considered by organizations (as discussed in the section below).

Key compliance requirements

The G7's AI Regulations set out the following 11 Guiding Principles and supplementary guidance:

- Risk identification and mitigation: Take appropriate measures throughout the development of Advanced AI systems, including prior to and throughout their deployment and placement on the market, to identify, evaluate, and mitigate risks across the AI lifecycle.[9] Testing should take place before deployment and before placement on the market, and should continue throughout the AI lifecycle.[10]

- Monitoring for risks and vulnerabilities: Organizations should monitor AI systems for vulnerabilities, incidents, emerging risks, and misuse after deployment, and take appropriate action to address these.[11] This includes, for example, facilitating third-party and user discovery and reporting of issues and vulnerabilities after deployment such as through bounty systems, contests, or prizes to incentivize the responsible disclosure of weaknesses. Appropriate documentation should be maintained, and reports on vulnerabilities should be accessible to a diverse set of stakeholders.[12]

- Accountability: Organizations should publicly report Advanced AI systems' capabilities, limitations and domains of appropriate and inappropriate use, to support ensuring sufficient transparency, with the aim of increasing accountability. This should include publishing transparency reports[13] containing meaningful information for all new significant releases of Advanced AI systems.[14]

- Information sharing: Organizations should work towards responsible information sharing and reporting of incidents.[15] Organizations should establish or join mechanisms to develop, advance, and adopt, where appropriate, shared standards, tools, mechanisms, and best practices across the AI lifecycle for ensuring the safety, security, and trustworthiness of Advanced AI systems.[16]

- AI governance: Organizations should develop, implement and disclose AI governance and risk management policies, grounded in a risk-based approach – including privacy policies, and mitigation measures, in particular for organizations developing Advanced AI systems.[17] This includes disclosing any appropriate privacy policies, including for personal data, user prompts and Advanced AI system outputs.[18]

- Security: Organizations should invest in and implement robust security controls, including physical security, cybersecurity and insider threat safeguards across the AI lifecycle. These controls may include securing model weights and algorithms, servers, and datasets such as through operational security measures for information security and appropriate cyber/physical access controls.[19] Organizations should also look to establish a robust insider threat detection program.[20]

- Authentication and provenance: Organizations should develop and deploy reliable content authentication and provenance mechanisms such as watermarking or other techniques to enable users to identify AI-generated content. Organizations should also develop tools or APIs to allow users to determine if particular content was created with their Advanced AI system, as well as other mechanisms such as labelling or disclaimers to enable users, where possible and appropriate, to know when they are interacting with an AI system.[21] The provenance data need not identify individual users. Content authentication mechanisms should be developed by organizations only where technically feasible and appropriate.[22]

- Research and development: Organizations should prioritize research to mitigate societal, safety and security risks and prioritize investment in effective mitigation tools.[23] Organizations should prioritize research on key areas such as upholding democratic values, respecting human rights, protecting children and vulnerable groups, safeguarding intellectual property rights and privacy, and avoiding harmful bias, misinformation and disinformation, and information manipulation.[24]

- Focus on challenges: Organizations should prioritize the development of Advanced AI systems to address global challenges such as climate change, global health and education. These efforts are undertaken in support of progress on the United Nations Sustainable Development Goals, and to encourage AI development for global benefit.[25] Organizations should prioritize responsible stewardship of trustworthy and human-centric AI and also support digital literacy initiatives that promote the education and training of the public, including students and workers.[26]

- Development of standards: Organizations should advance the development and adoption of international technical standards and best practices.[27] In particular, organizations are encouraged to work to develop interoperable international technical standards and frameworks to help users distinguish content generated by AI from non-AI generated content.[28]

- Data protection and IP: Organizations should implement appropriate protections for personal data and intellectual property, as well as transparency of training datasets.[29] Appropriate measures could include transparency, privacy-preserving training techniques, and/or testing and fine-tuning to ensure that systems do not divulge confidential or sensitive data.[30]

Regulators

The G7 intends to develop, in consultation with the OECD and other stakeholders, monitoring tools and mechanisms to help AI actors "stay accountable" in their compliance with the Guiding Principles and Code of Conduct.[31] This suggests that the G7's AI Regulations will therefore be "self-regulated" by the organizations and/or individuals to which they apply, but the position is not settled.

The G7's AI Regulations do not otherwise stipulate how the G7 nations should regulate the implementation of the Guiding Principles in their own jurisdictions.

Enforcement powers and penalties

As the G7's AI Regulations are not legally binding, they do not confer enforcement powers or give rise to any penalties for non-compliance. The G7 therefore relies on its members to implement the relevant Guiding Principles and give effect to the Code of Conduct. Notably, the Guiding Principles and Code of Conduct state that each G7 state has considerable discretion to implement the relevant AI Regulation uniquely in different ways and as each sees fit.[32]

1 See here.

2 See here.

3 See the G7 Leaders' Statement.

4 See G7 Joint Statement here.

5 See G7 Discussion Paper here.

6 Please see the OECD's AI Principles.

7 See the Guiding Principles (draft), page 1, paragraph 1.

8 See the Guiding Principles (draft), page 2, paragraph 1.

9 This includes testing measures such as "red-teaming" and traceability in relation to datasets, processes and decisions. See the Guiding Principles (draft), page 2, Principle 1.

10 See the Code of Conduct (draft), pages 2-3, Code 1. Relevant risks include: (i) chemical, biological, radiological and nuclear risks; (ii) offensive cyber capabilities; (iii) risks to health and/or safety; (iv) risks from models "self-replicating" themselves or training other models; (v) societal risks; (vi) threats to democratic values and human rights; and (vii) risks of creating a chain reaction.

11 See the Guiding Principles (draft), page 2, Principle 2.

12 See the Code of Conduct (draft), page 4, Code 2.

13 See the Code of Conduct (draft), page 4, Code 3. Transparency reports should include, for example: (i) details of the evaluations conducted for potential safety, security and societal risks; (ii) capacities of the model/system and significant limitations in performance that have implications for the domains of appropriate use; (iii) assessment of the AI system's effects and risks, such as harmful bias, discrimination and threats to the protection of privacy or personal data; and (iv) the results of "red-teaming."

14 See the Guiding Principles (draft), page 3, Principle 3.

15 See the Guiding Principles (draft), page 3, Principle 4.

16 See the Code of Conduct (draft), page 5, Code 4.

17 See the Guiding Principles (draft), page 3, Principle 5.

18 See the Code of Conduct (draft), pages 5-6, Code 5.

19 See the Guiding Principles (draft), pages 4, Principle 6.

20 See the Code of Conduct (draft), page 6, Code 6.

21 See the Guiding Principles (draft), page 4, Principle 7.

22 See the Code of Conduct (draft), pages 6-7, Code 7.

23 See the Guiding Principles (draft), page 4, Principle 8.

24 See the Code of Conduct (draft), page 7, Code 8.

25 See the Guiding Principles (draft), pages 4-5, Principle 9.

26 See the Code of Conduct (draft), pages 7-8, Code 9.

27 See the Guiding Principles (draft), page 5, Principle 10.

28 See the Code of Conduct (draft), page 8, Code 10.

29 See the Guiding Principles (draft), page 5, Principle 11.

30 See the Code of Conduct (draft), page 8, Code 11.

31 See the Guiding Principles (draft), page 1, paragraph 6 and the Code of Conduct (draft), Page 1.

32 See the Guiding Principles (draft), page 1, paragraph 5 and the Code of Conduct (draft), Page 1.

White & Case means the international legal practice comprising White & Case LLP, a New York State registered limited liability partnership, White & Case LLP, a limited liability partnership incorporated under English law and all other affiliated partnerships, companies and entities.

This article is prepared for the general information of interested persons. It is not, and does not attempt to be, comprehensive in nature. Due to the general nature of its content, it should not be regarded as legal advice.

© 2024 White & Case LLP